

Text Mining

Task 1 & 2

Xiang Li

12/3/2021

Task 1

I choose **My Doggie and I** by R.M. Ballantyne as the resource of my text analysis. **My Doggie And I** tells the story of John Mellon (almost a doctor in chapter one) and what happens to him after he meets a certain little canine. This story surrounds a child waif, a young woman, a young gentleman doctor, and an elderly lady. This tale unfolds the story of a bond that brings these unlikely friends together and merges their separate paths of life into one common path.

First, let us get a basic sense of most common words in the book.

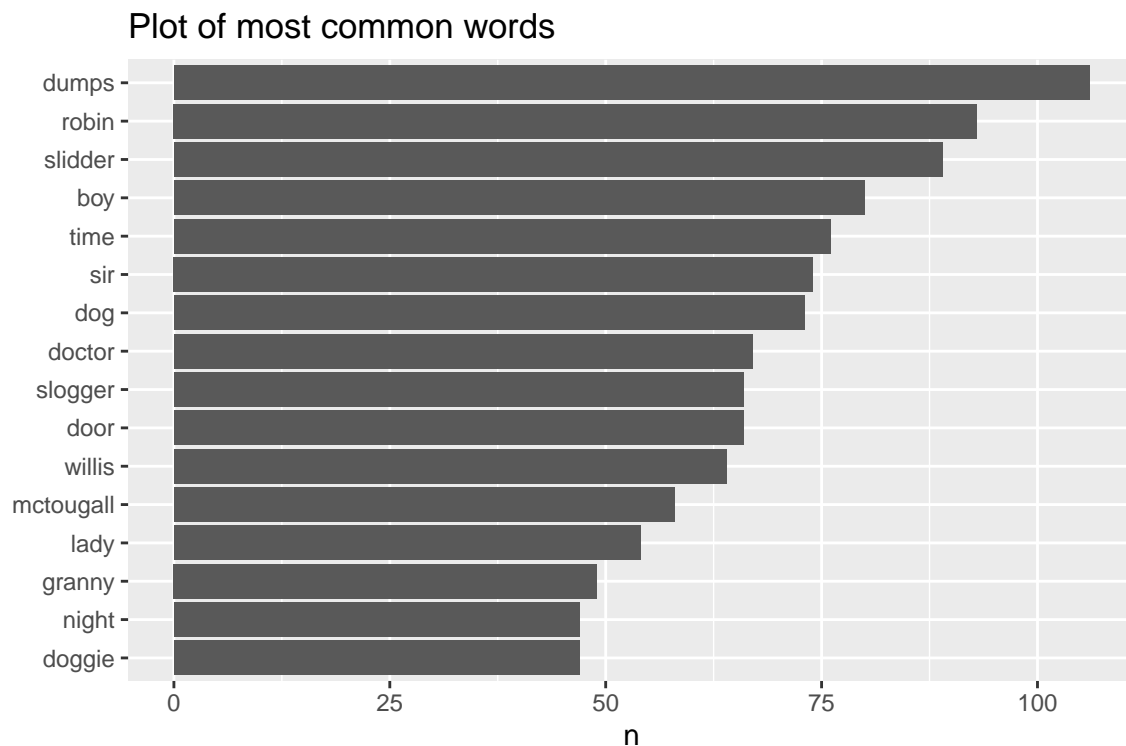


Figure 1: A visualization of the most common words

Task 2- Bag of Words Analysis

With several options for sentiment lexicons, we can use `inner_join()` to calculate the sentiment in different ways. Following the step in **Textmining in R**, I choose three general-purpose lexicons are *AFINN* from Finn Årup Nielsen, *bing* from Bing Liu and collaborators, and **nrc* from Saif Mohammad and Peter Turney. and generat three plot which showed an estimate of the net sentiment (positive - negative) in each chunk of the novel text for each sentiment lexicon.

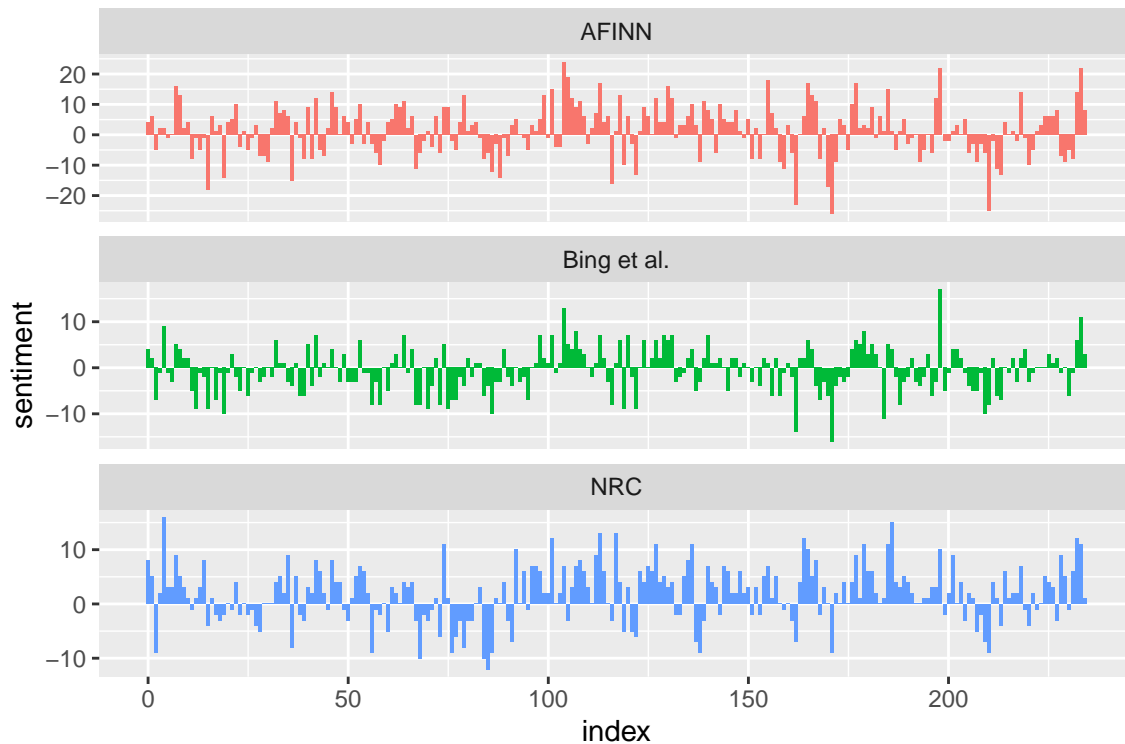
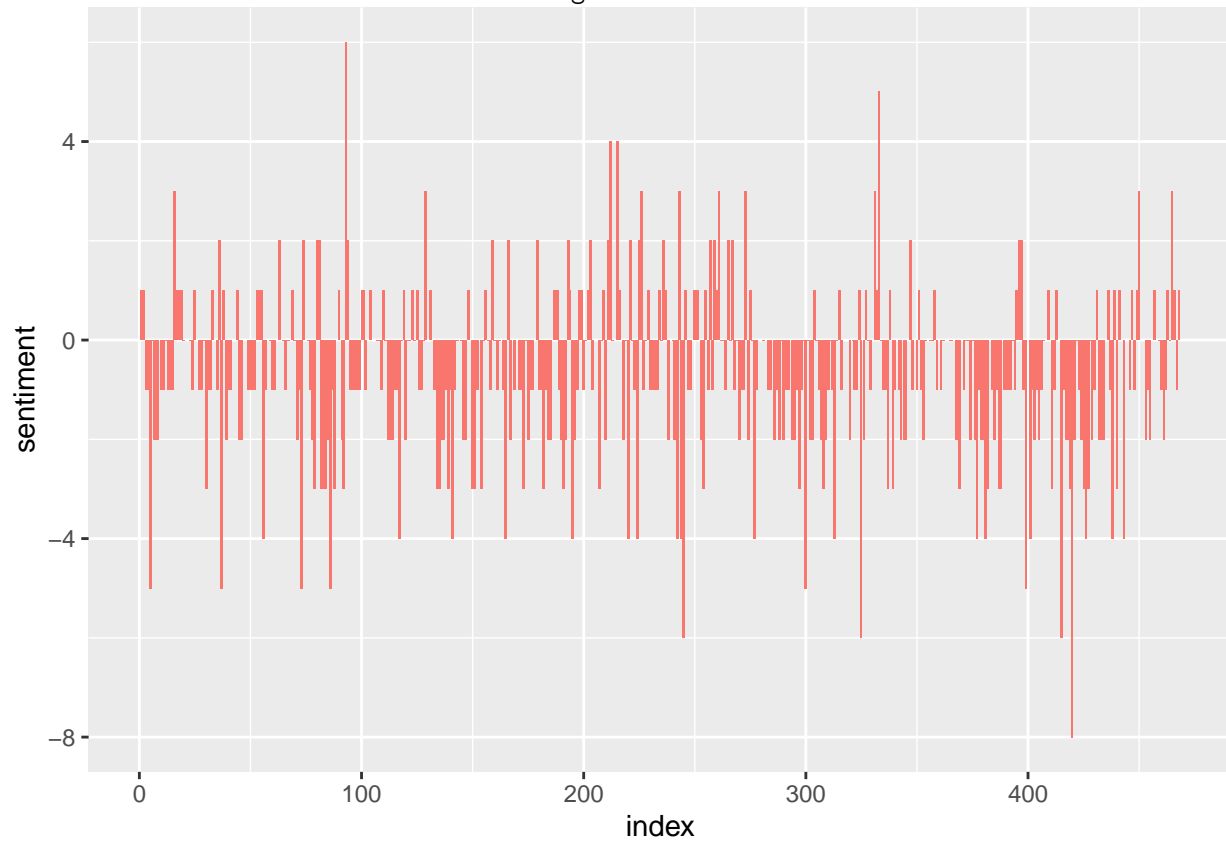


Figure 2: visualization of an estimate of the net sentiment (positive - negative) for each sentiment lexicon

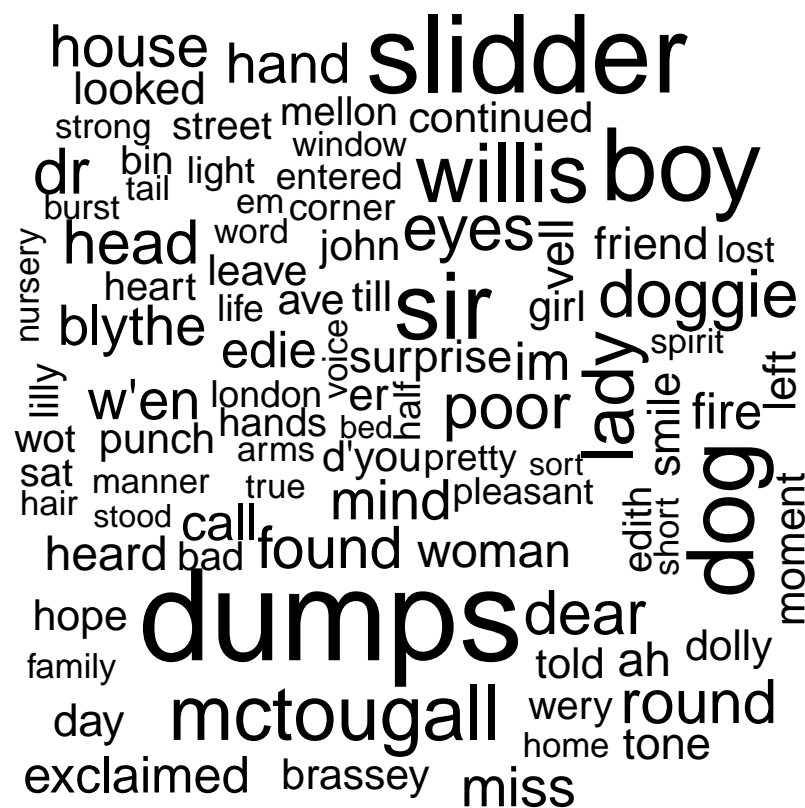
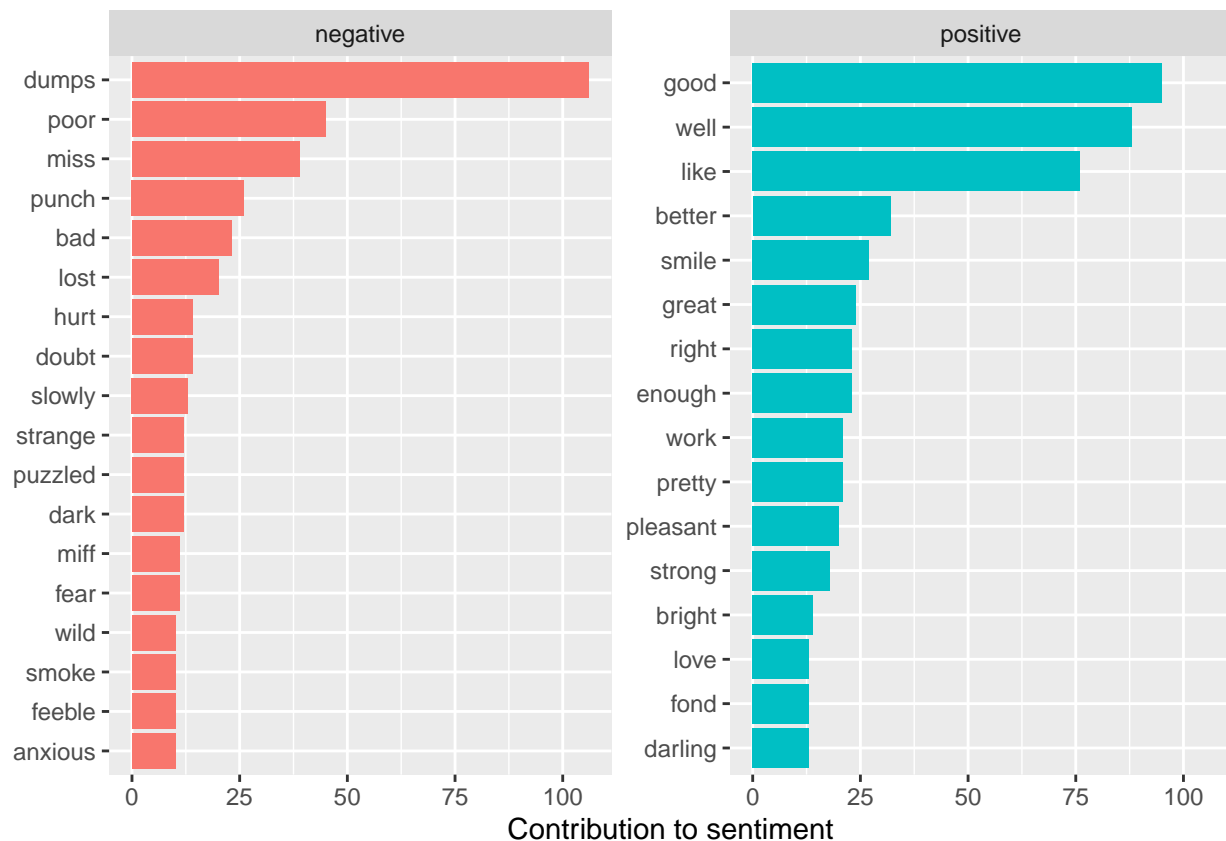
Discussion:

Task 2 Extra credit

The additional lexicons I found is that loughran. And here is the sentiment flow of the book.



Discussion:



negative



positive